BACKWARD DRIFT ESTIMATION WITH APPLICATION TO QUALITY LAYER ASSIGNMENT IN H.264/AVC BASED SCALABLE VIDEO CODING

Thomas Rusert and Jens-Rainer Ohm

Institute of Communications Engineering
RWTH Aachen University
52056 Aachen, Germany

ABSTRACT

We present an approach for accurate estimation of the reconstruction distortion in SNR scalable video coding with drift. Based on a linear model of predictive video coding, we derive an algorithm to quantify spatio-temporal drift properties subject to prediction structure and motion information. This allows for low-complex estimation of the reconstruction distortion on a per-block basis. The accuracy of the distortion estimation is experimentally verified. We then utilize the method for quality layer assignment within the framework of H.264/AVC scalable video coding (SVC), which is currently under standardization. The quality layers allow for bit stream truncation in a rate-distortion optimized sense. Compared to the quality layer assignment as implemented in the SVC test model, use of backward drift estimation allows for achieving equivalent coding efficiency with reduced complexity.

Index Terms— Error propagation, hierarchical B pictures, quality layers, SVC, H.264/AVC

1. INTRODUCTION

Motion compensated temporal filtering (MCTF) has proven to provide a robust basis for highly efficient scalable video coding. Hierarchical temporal prediction based on B pictures can be seen as a specific instantiation of MCTF. It is a fundamental element of H.264/AVC scalable video coding (SVC), which is currently jointly developed by ISO/IEC MPEG and ITU-T VCEG [1, 2]. The hierarchical B prediction structure enables effective attenuation of error drift, which is inevitably caused if a SVC bit stream is decoded at a bit rate lower than that used for operating the prediction loop at the encoder [3]. The propagation of an encoder-decoder mismatch at a particular spatio-temporal location in the reconstruction process is strictly related to the prediction correspondences, and is therefore strongly dependent on the motion information. In our previous work [4], we have developed a linear model of the prediction process under consideration of the coding control, based on which the impact of drift can be estimated. Using a heuristic approach to quantify the propagation properties, it was shown that consideration of the potentially remaining drift can allow for optimization of the bit allocation during encoding, and consequently improved coding efficiency.

In this paper, we propose an analytic approach for quantification of the spatio-temporal error propagation properties in predictive video coding. Using the presented backward drift estimation algorithm, accurate error propagation correspondences, considering the exact mode information and sub-pel accurate motion vectors, can be derived on a per-pixel basis. To reduce the computational complexity, the algorithm is generalized to derive correspondences on a per-block basis or a per-picture basis.

While in existing drift estimation approaches, such as [5, 6], the expected distortion is tracked subject to a-priori knowledge of quantization errors or quantization error probabilities, our approach derives generic correspondences between spatio-temporal regions, which can then later be utilized to determine the impact of individual quantization errors. This allows for accurate rate-distortion optimized bit allocation. Furthermore, other than [5, 6], our approach is capable of coping with error correlations caused by error propagation over different paths, which can frequently occur within hierarchical B prediction structures. Additionally, sub-pel interpolation can be accurately considered.

We verify the accuracy of our drift estimation method by predicting the reconstruction distortion based on the per-pixel quantization error. It is then applied to quality layer assignment in SVC. Use of quality layers allows for bit stream truncation in a rate-distortion optimized sense [2, 7]. We show that compared to the quality layer assignment as implemented in the SVC test model, use of backward distortion estimation allows for achieving equivalent coding efficiency with reduced complexity. The approach is similarly applicable to other bit allocation problems in predictive video coding schemes with drift, such as [4, 8].

The paper is organized as follows. The key elements of the investigated system are outlined in Sec. 2. In Sec. 3, we review the formulation of our linear distortion model for predictive video coding with drift. The drift estimation algorithm is developed in Sec. 4. In Sec. 5 we provide experimental verification of our approach, and utilize the method for optimized quality layer assignment in SVC. Sec. 6 concludes the paper.

2. INVESTIGATED SYSTEM

2.1. Hierarchical B Pictures

In Fig. 1a, a temporal prediction structure with hierarchical B pictures and $T = 3$ levels is illustrated. f_z^t denotes the picture at temporal position z and temporal resolution t. Here, $t = 0$ corresponds to the coarsest temporal resolution, and $t = T - 1$ corresponds to the finest resolution. The arrows indicate the prediction dependencies, e.g. picture f_1^2 is predicted by a bi-directional motion compensated reference picture generated from pictures f_0^0 and f_2^1. In [1, 2], flexible reference picture positions can be signaled for prediction. However, the SVC test model implements the fundamental dyadic decomposition structure according to Fig. 1a, where a picture f_z^t is predicted from the two nearest neighboring pictures associated with a temporal resolution less than t.

This work was funded by Deutsche Forschungsgemeinschaft (DFG) under contract OH 50/11-1.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
APR 2007		**00-00-2007 to 00-00-2007**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Backward Drift Estimation with Application to Quality Layer Assignment in H.264/AVC Based Scalable Video Coding	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Institute of Communications Engineering,RWTH Aachen University,52056 Aachen, Germany,**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002013. Proceedings of the 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Held in Honolulu, Hawaii on April 15-20, 2007. Government or Federal Rights

14. ABSTRACT
We present an approach for accurate estimation of the reconstruction distortion in SNR scalable video coding with drift. Based on a linear model of predictive video coding, we derive an algorithm to quantify spatio-temporal drift properties subject to prediction structure and motion information. This allows for low-complex estimation of the reconstruction distortion on a per-block basis. The accuracy of the distortion estimation is experimentally verified. We then utilize the method for quality layer assignment within the framework of H.264/AVC scalable video coding (SVC), which is currently under standardization. The quality layers allow for bit stream truncation in a rate-distortion optimized sense. Compared to the quality layer assignment as implemented in the SVC test model, use of backward drift estimation allows for achieving equivalent coding efficiency with reduced complexity.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**4**	

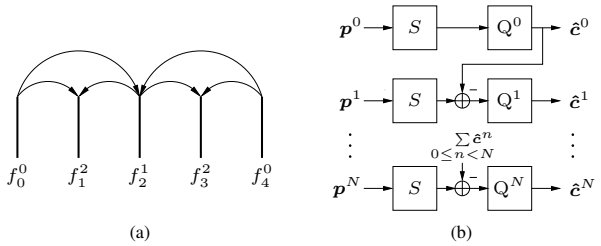

Fig. 1. (a) Hierarchical B prediction structure with $T = 3$ levels. (b) Progressive refinement quantization with N FGS layers.

2.2. Progressive Refinement Quantization

SNR scalability in SVC is enabled by utilization of either enhancement slices (CGS/MGS), or progressive refinement slices (FGS) [2]. The approaches differ in their respective quantization and coding schemes, representing different trade-offs between complexity and granularity of scalability. While our presented distortion model holds equally for either scheme, we employ FGS coding in this paper.

The basic FGS quantization principle is illustrated in Fig. 1b. Here, $\boldsymbol{p}^n$ denotes the unquantized prediction residual information, $0 \le n \le N$, and $\hat{\boldsymbol{c}}^n$ denotes the quantized transform coefficients of the nth FGS layer, where $n = 0$ corresponds to the quality base layer. Note that when multiple prediction loops are operated at the encoder, the input signals $\boldsymbol{p}^n$ may be different from each other.

The forward and backward spatial transform operations are denoted as $\boldsymbol{S}$ and $\boldsymbol{S}^{-1}$, respectively, and the quantization stage associated with layer n is represented by Q^n. For each FGS layer, quantization is performed on the refinement information relative to the preceding layers. To allow for progressive refinement, the quantizer step sizes must be monotonically decreasing with n. In SVC, the step size is halved with each FGS layer.

SVC enables fine granular SNR scalability on the bit stream level by truncation of network abstraction layer units containing progressive refinement information (PR-NALUs). To allow for low-complex rate-distortion optimized bit stream truncation, the relative importance of each of the respective NALUs can be signaled in the NALU headers. This concept is denoted as quality layers [7].

3. MODEL-BASED ESTIMATION OF THE RECONSTRUCTION DISTORTION

In the following, we review the linear formulation of the prediction operations introduced in [4, 8], and derive a linear distortion model for an SVC codec. While throughout this paper we assume that a single prediction loop is operated at the highest FGS layer [3], i.e. $\boldsymbol{p}^n = \boldsymbol{p}^N = \boldsymbol{p}, \forall n$, the model can be generalized for the case of multiple prediction loops according to [4]. Fixed motion information is assumed, and non-linear effects (rounding, clipping, deblocking) are neglected. We denote $\boldsymbol{x}$ as an $L \times 1$ vector comprising the L original (unquantized) samples of a dependently coded subset of the input video sequence (e.g., see Fig. 1a). The linear prediction operations are expressed as follows.

$$\boldsymbol{p} = \boldsymbol{x} - \boldsymbol{M}\hat{\boldsymbol{x}}^{\mathrm{enc}} - \boldsymbol{I}\hat{\boldsymbol{x}}^{\mathrm{bl}} - \boldsymbol{k} \tag{1}$$

Here, the unquantized prediction residual is represented as an $L \times 1$ vector $\boldsymbol{p}$. The $L \times L$ matrices $\boldsymbol{M}$ and $\boldsymbol{I}$ express the motion-compensated temporal prediction and the directional intra prediction, respectively. $\hat{\boldsymbol{x}}^{\mathrm{enc}}$, $\hat{\boldsymbol{x}}^{\mathrm{bl}}$ and $\boldsymbol{k}$ are $L \times 1$ vectors, where $\hat{\boldsymbol{x}}^{\mathrm{enc}}$ denotes the prediction reference used within the encoder prediction loop, and $\hat{\boldsymbol{x}}^{\mathrm{bl}}$ represents the sequence reconstructed from the quality base layer without FGS refinement. $\boldsymbol{k}$ represents a static prediction such as intra DC prediction. Note that $\hat{\boldsymbol{x}}^{\mathrm{bl}}$ is used as intra prediction reference [2] since it is guaranteed to be available to the decoder regardless of the amount of FGS refinement. Thus, any intra prediction drift is avoided. Assuming the samples in $\boldsymbol{x}$ are arranged in macroblock coding order, $\boldsymbol{M}$ and $\boldsymbol{I}$ are strictly lower triangular matrices. Moreover, since intra prediction is constrained [2], $I_{i,j} = 0$ if not both i and j are indices of intra coded pixels.

The spatial forward transform, quantization, and backward transform processes can be formulated as

$$\begin{aligned} \boldsymbol{c} &= \boldsymbol{S}\boldsymbol{p}, & (2)\\ \hat{\boldsymbol{p}} &= \boldsymbol{S}^{-1}(\boldsymbol{c}+\boldsymbol{q}) \\ &= \underbrace{\boldsymbol{S}^{-1}\boldsymbol{c}}_{\boldsymbol{p}} + \underbrace{\boldsymbol{S}^{-1}\boldsymbol{q}}_{\boldsymbol{e}}. & (3) \end{aligned}$$

Here, $\boldsymbol{S}$ and $\boldsymbol{S}^{-1}$ are $L \times L$ matrices expressing the spatial forward and backward transform operations, respectively, generating the $L \times 1$ vector of transform coefficients, $\boldsymbol{c}$. Quantization is represented by addition of an $L \times 1$ random vector $\boldsymbol{q}$, which depends on the decoded bit rate. $\hat{\boldsymbol{p}}$ and $\boldsymbol{e}$ denote the reconstructed residual signal and the quantization error after backward transform, respectively.

The reconstruction is generated based on the prediction references available to the decoder. From Eq. (3) and Eq. (1), it follows

$$\begin{aligned} \hat{\boldsymbol{x}} &= \hat{\boldsymbol{p}} + \boldsymbol{M}\hat{\boldsymbol{x}} + \boldsymbol{I}\hat{\boldsymbol{x}}^{\mathrm{bl}} + \boldsymbol{k} \\ &= \boldsymbol{x} + \boldsymbol{M}\left(\hat{\boldsymbol{x}} - \hat{\boldsymbol{x}}^{\mathrm{enc}}\right) + \boldsymbol{e}. & (4) \end{aligned}$$

Decoding the full bit stream including all FGS layers equivalents a classical drift-free prediction scheme with $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}^{\mathrm{enc}}$ and

$$\hat{\boldsymbol{x}}^{\mathrm{enc}} - \boldsymbol{x} = \boldsymbol{e}^{\mathrm{enc}}. \tag{5}$$

For reconstruction with quantization error $\boldsymbol{e}^{\mathrm{dec}}$, substituting Eq. (5) into Eq. (4) yields [4]

$$\hat{\boldsymbol{x}}^{\mathrm{dec}} - \boldsymbol{x} = \boldsymbol{e}^{\mathrm{dec}} + \boldsymbol{B}\left(\boldsymbol{e}^{\mathrm{dec}} - \boldsymbol{e}^{\mathrm{enc}}\right), \tag{6}$$

$$\text{with} \quad \boldsymbol{B} + \boldsymbol{1} = (\boldsymbol{1} - \boldsymbol{M})^{-1}. \tag{7}$$

Here, $\boldsymbol{1}$ is the $L \times L$ identity matrix, and $\boldsymbol{B}$ is a strictly lower triangular $L \times L$ matrix.

Since there is no inter prediction within a picture, we observe that $B_{i,j} = 0$ if both $i, j \in f_z$. Furthermore, we assume that quantization errors in different pictures are uncorrelated, i.e. $\mathrm{E}[e_i e_j] = 0$ if the indices i and j do not belong to the same picture, with $\mathrm{E}[\cdot]$ denoting the expectation. For samples within a given picture f_z, we further assume that

$$\mathrm{E}[\sum_{i=0}^{L-1}\sum_{\substack{j=0\\ j\neq i}}^{L-1}\left(e_i^{\mathrm{dec}} - e_i^{\mathrm{enc}}\right)\left(e_j^{\mathrm{dec}} - e_j^{\mathrm{enc}}\right)\sum_{k\in f_z} B_{k,i}B_{k,j}] = 0. \tag{8}$$

This is reasonable for high bit rates, where quantization errors can be assumed to be uncorrelated. It is also reasonable for homogeneous full-pel motion, where neighboring quantization errors should not interact during reconstruction, i.e. $B_{k,i}B_{k,j} = 0$. With these assumptions, we can formulate the expected quadratic distortion of picture f_z as follows.

$$\begin{aligned} E[D_{f_z}] &= E[\sum_{i\in f_z}\left(\hat{x}_i^{\mathrm{dec}} - x_i\right)^2] \\ &= \sum_{i\in f_z}\left(\left(e_i^{\mathrm{dec}}\right)^2 + \sum_{\substack{j=0\\ j\notin f_z}}^{L-1} B_{i,j}^2\left(e_j^{\mathrm{dec}} - e_j^{\mathrm{enc}}\right)^2\right) & (9) \end{aligned}$$

It can be seen that the squared matrix elements $B_{i,j}^2$ determine the expected distortion contribution to the reconstruction at position i, caused by a drift term introduced at position j. Note that as of the generic formulation of the model, the elements $B_{i,j}^2$ can reflect any temporal prediction structure. Particularly, sub-pel interpolation is seamlessly integrated, and for hierarchical prediction structures, Fig. 1a, the case where multiple correspondences over different paths exist between i and j is accurately represented.

4. BACKWARD DRIFT ESTIMATION ALGORITHM

Considering the triangularity of $\boldsymbol{M}$ and $\boldsymbol{B}$, we derive an algorithm to determine the correspondence factors $B_{i,j}^2$ as follows. From Eq. (7) it can be shown that $\boldsymbol{B} = (\boldsymbol{B} + \boldsymbol{1})\,\boldsymbol{M}$. Equivalently, we write

$$B_{i,j}(l) \quad = \quad M_{i,j} + \sum_{k=L-1}^{l+1} B_{i,k} M_{k,j},\ \forall i > j, \qquad (10)$$

such that $B_{i,j} = B_{i,j}(-1)$. Note that the sum index k is counted in descending order. This definition is convenient in the derivation of the algorithm below. To allow for both pixel based and block based derivation of correspondence factors, we formulate the expectation of the correspondence between two blocks $\mathcal{I}$ and $\mathcal{J}$, with $A = |\mathcal{I}| = |\mathcal{J}|$ the number of pixels per block.

$$\begin{aligned} B_{\mathcal{I}\mathcal{J}}(l) \;&:=\; E_{\substack{i\in\mathcal{I}\\ j\in\mathcal{J}}}[B_{i,j}(l)] \\ &=\; \frac{1}{A^2}\sum_{\substack{i\in\mathcal{I}\\ j\in\mathcal{J}}} M_{i,j} + \frac{1}{A}\sum_{j\in\mathcal{J}}\sum_{k=L-1}^{l+1} B_{\mathcal{I}\mathcal{K}} M_{k,j}, \qquad (11) \\ BS_{\mathcal{I}\mathcal{J}}(l) \;&:=\; E_{\substack{i\in\mathcal{I}\\ j\in\mathcal{J}}}[B_{i,j}^2(l)] \\ &=\; \frac{1}{A^2} E[\sum_{\substack{i\in\mathcal{I}\\ j\in\mathcal{J}}} \left(M_{i,j} + \sum_{k=L-1}^{l+1} B_{i,k} M_{k,j} \right)^2] \\ &=\; \frac{1}{A^2}\sum_{\substack{i\in\mathcal{I}\\ j\in\mathcal{J}}} \Bigg(M_{i,j}^2 + \sum_{k=L-1}^{l+1} BS_{\mathcal{I}\mathcal{K}} M_{k,j}^2 + \qquad (12) \\ &\quad 2\sum_{k=L-1}^{l+1} M_{k,j} E[B_{i,k}\left(M_{i,j} + \sum_{m=L-1}^{k+1} B_{i,m} M_{m,j} \right)] \Bigg) \end{aligned}$$

Here, $B_{\mathcal{I}\mathcal{J}} = B_{\mathcal{I}\mathcal{J}}(-1)$ and $BS_{\mathcal{I}\mathcal{J}} = BS_{\mathcal{I}\mathcal{J}}(-1)$. With Eq. (10), we write the last term in Eq. (12) as

$$\begin{aligned} &\frac{2}{A^2}\sum_{\substack{i\in\mathcal{I}\\ j\in\mathcal{J}}}\sum_{k=L-1}^{l+1} M_{k,j} E[B_{i,k} B_{i,j}(k)] \\ =\; &\frac{2}{A}\sum_{j\in\mathcal{J}}\sum_{k=L-1}^{l+1} M_{k,j}\Big(B_{\mathcal{I}\mathcal{K}} B_{\mathcal{I}\mathcal{J}}(k) + \\ &\rho\sqrt{(BS_{\mathcal{I}\mathcal{K}} - B_{\mathcal{I}\mathcal{K}}^2)\left(BS_{\mathcal{I}\mathcal{J}}(k) - B_{\mathcal{I}\mathcal{J}}^2(k)\right)}\Big), \qquad (13) \end{aligned}$$

where ρ accounts for cross-correlations between the elements contributing to $B_{\mathcal{I}\mathcal{K}}$ and $B_{\mathcal{I}\mathcal{J}}(k)$. Finally, from Eq. (11) – Eq. (13), we derive the following algorithm.

initialize $B_{\mathcal{I}\mathcal{J}} = 0$, $BS_{\mathcal{I}\mathcal{J}} = 0, \forall \mathcal{I}, \mathcal{J}$
scan sequence in reverse coding order, $\forall \mathcal{K}$
  scan pixels $k \in \mathcal{K}$ *(direct establishing)*
    scan $j \in \mathcal{J}, \forall \mathcal{J}$, *such that* $M_{k,j} \neq 0$
      $B_{\mathcal{K}\mathcal{J}} \leftarrow B_{\mathcal{K}\mathcal{J}} + \frac{1}{A^2} M_{k,j}$
      $BS_{\mathcal{K}\mathcal{J}} \leftarrow BS_{\mathcal{K}\mathcal{J}} + \frac{1}{A^2} M_{k,j}^2$
  scan $\forall \mathcal{I}$, *such that* $BS_{\mathcal{I}\mathcal{K}} \neq 0$
    scan pixels $k \in \mathcal{K}$ *(indirect establishing)*
      $\theta \leftarrow B_{\mathcal{I}\mathcal{K}} B_{\mathcal{I}\mathcal{J}} + \rho\sqrt{(BS_{\mathcal{I}\mathcal{K}} - B_{\mathcal{I}\mathcal{K}}^2)\left(BS_{\mathcal{I}\mathcal{J}} - B_{\mathcal{I}\mathcal{J}}^2\right)}$
      scan $j \in \mathcal{J}, \forall \mathcal{J}$, *such that* $M_{k,j} \neq 0$
        $B_{\mathcal{I}\mathcal{J}} \leftarrow B_{\mathcal{I}\mathcal{J}} + \frac{1}{A} B_{\mathcal{I}\mathcal{K}} M_{k,j}$
        $BS_{\mathcal{I}\mathcal{J}} \leftarrow BS_{\mathcal{I}\mathcal{J}} + \frac{1}{A} BS_{\mathcal{I}\mathcal{K}} M_{k,j}^2 + \frac{2}{A} M_{k,j}\theta$

The video sequence is scanned in backward coding order. For each motion compensated temporal prediction $M_{k,j}$, the respective elements $B_{\mathcal{K}\mathcal{J}}$, $BS_{\mathcal{K}\mathcal{J}}$ are updated (*direct establishing*). Furthermore, for each existing correspondence $BS_{\mathcal{I}\mathcal{K}}$ originating from $\mathcal{K}$, the elements $B_{\mathcal{I}\mathcal{J}}$, $BS_{\mathcal{I}\mathcal{J}}$ are updated (*indirect establishing*). For $A = 1$, the algorithm accurately calculates the results of Eq. (11) – Eq. (13), with $B_{i,j}^2 = BS_{\mathcal{I}\mathcal{J}}$. For $A > 1$, $BS_{\mathcal{I}\mathcal{J}}$ is used as an approximation for $B_{i,j}^2$.

The computational complexity of the algorithm depends on the number of *establishing* steps to be performed. The number of *indirect establishing* steps depends on the respective number of existing correspondences $BS_{\mathcal{I}\mathcal{K}}$, which can be roughly expected to scale with $1/A$. For picture based derivation, the number of correspondences equals the number of dependent pictures. For a T-level B prediction structure with $T > 2$, it can be shown that the average number of pictures depending on a B picture is less than $T - 1$. Hence, since each B picture requires one *direct establishing step*, at most T *establishing* steps are performed in average. For a given k, the inner loops over j for $M_{k,j} \neq 0$ represent the prediction dependencies including interpolated pixels. For picture based derivation, the contributions of the individual interpolation taps can be summed-up instead of processing each tap separately. It can therefore be assumed that the complexity of each *establishing* step is $c < 1$, where $c = 1$ represents the complexity of the motion compensation operations for the respective picture. It follows that the total complexity of the algorithms is at most cT times the complexity of the motion compensation operations used for reconstruction of the video sequence.

5. EXPERIMENTAL RESULTS

In our experiments, we use the SVC test model JSVM-6 [3] with hierarchical B pictures. We use eight QCIF 15 Hz test sequences with different characteristics, and encode with $T = 5$ temporal levels and $N = 2$ FGS layers.

In the first experiment, we verify the accuracy of the distortion model and the drift estimation algorithm . We use two different block sizes, 4×4 and 176×144 (picture size), to quantify the drift correspondences. We extract each bit stream at 11 equally distributed target bit rates, and determine the quantization errors, $\boldsymbol{e}^{\mathrm{dec}}$. $\boldsymbol{e}^{\mathrm{enc}}$ is obtained from the non-truncated bit stream. We then use Eq. (9) to estimate the reconstruction distortion for each picture, and based on that, we calculate the estimated mean PSNR over each of the sequences. We thus obtain estimated PSNR measures for each of the extracted bit rates without performing the actual reconstruction.

Table 1 depicts the mean and maximum absolute differences of the estimated measures as compared to the true PSNR values obtained after decoding, for different values of ρ, with $\Delta = \mathrm{PSNR}_{\mathrm{est}} -$

	4×4			176×144	
ρ	mean $\lvert\Delta\rvert$	max $\lvert\Delta\rvert$	ρ	mean $\lvert\Delta\rvert$	max $\lvert\Delta\rvert$
0.0	0.350	1.392	0.0	0.376	1.452
0.2	0.259	1.091	0.1	0.315	1.233
0.4	0.191	0.725	0.2	0.267	1.004
0.6	0.233	0.526	0.3	0.233	0.753
0.8	0.414	0.899	0.4	0.245	0.577
1.0	0.695	1.347	0.5	0.305	0.753

Table 1. Mean and maximum absolute PSNR estimation error [dB].

$\text{PSNR}_{\text{true}}$. It can be seen that a mean absolute estimation error of about $0.2\,\text{dB}$ over all tested sequences can be achieved. Although the minimal achievable estimation errors are slightly lower for the case of 4×4 block based drift estimation, the results are very similar as for picture based estimation. We also observed similar results for pixel based derivation. This indicates that the distortion estimation error is primarily caused by inaccurate approximations used in the distortion model, see Sec. 3. We conclude that for picture based optimization of the bit allocation, it is sufficiently accurate to derive the drift correspondences on a per-picture basis. Derivation on smaller block bases will be advantageous for more localized bit allocation, such as in [4, 8].

We now use our distortion estimation technique for quality layer assignment. After encoding a sequence, the backward drift estimation algorithm is performed on a per-picture basis with $\rho = 0.3$. We then extract the quantization errors $\boldsymbol{e}^{\text{enc}}$, $\boldsymbol{e}^{\text{dec}}$ corresponding to the base layer and each of the FGS layers. Based on that, starting with the highest FGS layer, we use the distortion model to compute the estimated PSNR. Then the expected decrease ΔD_z in PSNR is calculated for the least significant remaining PR-NALU of each picture f_z, with ΔR_z the bit budget of that NALU. The PR-NALU with the lowest rate-distortion slope $\Delta D_z / \Delta R_z$ is assigned the least significant quality layer. Starting from the now estimated PSNR, the algorithm is repeated until all PR-NALUs are assigned a quality layer. The obtained quality information is possibly merged according to [7], such that the maximum number of quality layers defined in SVC is not exceeded.

The resulting coding performance of the scalable bit stream after quality layer assignment is exemplarily depicted in Fig. 2 . It can be seen that compared to the quality layer assignment in JSVM-6, our approach provides equivalent coding gain. Furthermore, while the quality layer assignment in JSVM-6 requires $2NT$ decoding passes to establish the required rate-distortion dependencies [3], our algorithm is less complex, requiring the equivalent of cT decoding passes, with $c < 1$.

6. CONCLUSION

We have presented a model-based method for drift estimation in SNR scalable video coding. The accuracy of the prediction has been experimentally verified. The method has then been utilized for quality layer assignment within the framework of H.264/AVC scalable video coding. For the case of single-loop FGS coding, compared to the quality layer assignment method in the SVC test model, our approach achieves equivalent coding efficiency with reduced computational complexity.

The presented distortion estimation approach can also be utilized for applications requiring consideration of spatially localized drift properties, such as macroblock-based bit allocation. As of its generality, the algorithm can principally be used for estimation of error propagation in any lifting-based MCTF system.

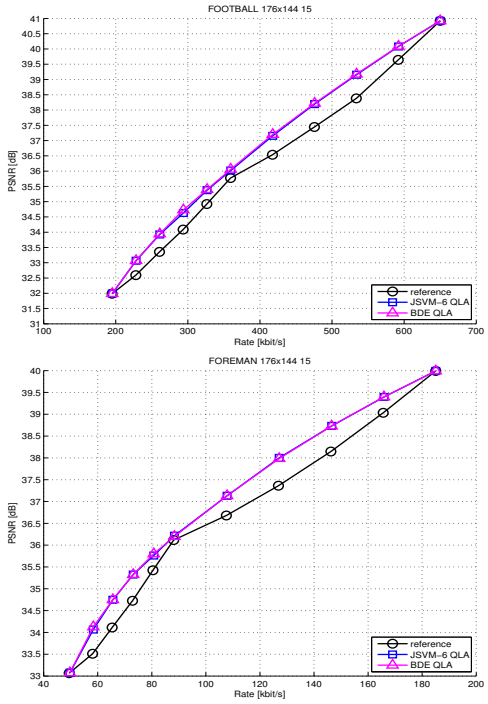

Fig. 2. Simulation results for FGS coding with quality layer assignment (QLA) based on backward drift estimation (BDE).

7. REFERENCES

[1] *ITU-T Rec. H.264 & ISO/IEC 14496-10 AVC: Advanced Video Coding for Generic Audiovisual Services*, version 3: 2005.

[2] J. Reichel, H. Schwarz, and M. Wien, "Scalable video coding – joint draft 6," Doc. JVT-S201, Joint Video Team (JVT), 19th Meeting, Geneva, Switzerland, Apr. 2006.

[3] J. Reichel, H. Schwarz, and M. Wien, "Joint scalable video model JSVM-6," Doc. JVT-S202, Joint Video Team (JVT), 19th Meeting, Geneva, Switzerland, Apr. 2006.

[4] T. Rusert and J.-R. Ohm, "Macroblock based bit allocation for SNR scalable video coding with hierarchical B pictures," in *Proc. IEEE Int. Conference on Image Processing ICIP '06*, Atlanta, GA, USA, Oct. 2006.

[5] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 966–976, June 2000.

[6] A. Leontaris and P. C. Cosman, "Drift-resistant SNR scalable video coding," *IEEE Trans. on Image Processing*, vol. 15, no. 8, pp. 2191–2197, Aug. 2006.

[7] I. Amonou, N. Cammas, S. Kervadec, and S. Pateux, "Optimized rate-distortion extraction with quality layers," in *Proc. IEEE Int. Conference on Image Processing ICIP '06*, Atlanta, GA, USA, Oct. 2006.

[8] T. Rusert, M. Spiertz, and J.-R. Ohm, "H.264/AVC compatible scalable multiple description video coding with RD optimization," in *Proc. IEEE Int. Workshop on Intelligent Signal Processing and Communication Systems ISPACS '06*, Yonago, Japan, Dec. 2006.